NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

MAXIMIZING WEAPON SYSTEM AVAILABILITY WITH A MULTI-ECHELON SUPPLY NETWORK

by

Brennan J. Kemper

June 2014

Thesis Co-Advisors:	Emily M. Craparo
	Javier Salmeron
Second Reader:	Walter DeGrange

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2014	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE **MAXIMIZING WEAPON SYSTEM AVAILABILITY WITH A MULTI-ECHELON SUPPLY NETWORK**			**5. FUNDING NUMBERS**
6. AUTHOR(S) Brennan J. Kemper			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB protocol number ____N/A____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Weapon systems are comprised of parts that are subject to random failures. When a part fails, it must be replaced by an operable part that is provided by a supply network that supports the system. Supply networks consist of many locations where spare parts are held, known as "echelons." Examples include depots, fulfillment centers, and customers. When many identical weapon systems operate in parallel and rely on a multi-echelon supply network for replacement parts, decision makers must choose where and how to invest their resources into the purchase of spare parts. This thesis uses stochastic optimization to leverage those decisions in order to maximize the expected number of time periods a set of weapon systems are available for use. Specifically, we develop a model that determines the optimal stock levels of spare parts to store at each echelon of the supply network. The formulation integrates part failure uncertainty, transit times, and monetary constraints. Model outputs also provide decision makers with a clear estimate of marginal availability gains for each dollar invested in purchasing spare parts.

14. SUBJECT TERMS availability, multi-echelon, network, optimization, repairable part, stochastic, stock level, supply.			**15. NUMBER OF PAGES** 71
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

MAXIMIZING WEAPON SYSTEM AVAILABILITY WITH A MULTI-ECHELON SUPPLY NETWORK

Brennan J. Kemper
Lieutenant Commander, United States Navy
B.S., Duquesne University, 2002

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2014

Author: Brennan J. Kemper

Approved by: Emily M. Craparo
Thesis Co-Advisor

Javier Salmeron
Thesis Co-Advisor

Walter DeGrange
Second Reader

Robert F. Dell
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Weapon systems are comprised of parts that are subject to random failures. When a part fails, it must be replaced by an operable part that is provided by a supply network that supports the system. Supply networks consist of many locations where spare parts are held, known as “echelons.” Examples include depots, fulfillment centers, and customers. When many identical weapon systems operate in parallel and rely on a multi-echelon supply network for replacement parts, decision makers must choose where and how to invest their resources into the purchase of spare parts. This thesis uses stochastic optimization to leverage those decisions in order to maximize the expected number of time periods a set of weapon systems are available for use. Specifically, we develop a model that determines the optimal stock levels of spare parts to store at each echelon of the supply network. The formulation integrates part failure uncertainty, transit times, and monetary constraints. Model outputs also provide decision makers with a clear estimate of marginal availability gains for each dollar invested in purchasing spare parts.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Weapon systems are comprised of parts that are subject to random failures. When a part fails, it must be replaced by an operable part that is provided by a supply network that supports the system. Supply networks consist of many locations where spare parts are held, known as "echelons." Examples include depots, fulfillment centers, and customers. If the supply network does not provide replacement parts in a timely manner, weapon system availability suffers. When many identical weapon systems operate in parallel and rely on a single supply network for replacement parts, decision makers must choose stock levels—the most efficient quantities of spare parts to stock at each echelon in a supply network, given a limited budget, so that system availability is maximized.

This thesis develops a stochastic optimization model that prescribes optimal investments in spare parts for a weapon system throughout a multi-echelon supply network. This model also provides decision makers with an estimate of expected system availability given such investments. Our formulation models the dynamic interactions of random part failures, transit times, and monetary constraints, all of which impact the supply network's ability to maintain weapon system availability.

Our model confronts the above problem in two steps. First, we incorporate uncertainty into our formulation by randomly generating scenarios of time-based part failures, which become input parameters for our stock level calculations. Second, we model the evolution of our inventory levels over time and incur a penalty in our objective function if a system is unavailable (i.e., if the supply network does not provide replacement parts in a timely manner for that system). Model outputs provide decision makers with optimal stock levels and a clear estimate of marginal availability gains they can expect for each dollar invested in purchasing spare parts.

We implement our formulation and show that as supply network topology becomes larger and more complex, more investments in spare parts are needed to ensure high rates of weapon system availability. Also, we find that marginal availability increases faster with additional investments in retail budget than in wholesale budget.

Moreover, we show that in some cases, we may obtain similar availability for different investment levels. For our specific example, a decision maker can expect 96 percent availability by optimally spending any budget between \$80 and \$95 on the procurement of spare parts for the system.

Additionally, we examine the variation in our model's output due to uncertainty. We show that increasing the number of part failure scenarios over which we optimize increases both solution quality and the amount of time required to find an optimal solution. We conclude that it is important to balance both aspects. For our particular example, we illustrate that modeling 25 part-failure scenarios per replication gives us results that reasonably balance objective value precision and solve time. Furthermore, we show that as the number of part-failure scenarios modeled increases, our model converges towards an optimal solution for our wholesale and retail stock levels.

Lastly, we examine restriction and relaxation techniques that decrease the time required to find an optimal solution. We achieve moderate success with these techniques but conclude that further research in this area is needed.

ACKNOWLEDGMENTS

I would like to extend my sincerest appreciation to **Professor Emily Craparo**. She is a truly dedicated professional and an absolutely wonderful person. Without her brilliance, patience, and guidance, I would not have been able to complete this thesis. I would also like to thank **Professor Javier Salmeron** for teaching me the concepts of linear programming and helping me shape the direction of my research efforts. My special thanks to **CDR Walt DeGrange**, who served as an outstanding program officer and mentor. Lastly, I would like to thank the operations research faculty at the Naval Postgraduate School. They have given me a skill set that I will carry for the rest of my life.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

1. Problem Definition

Weapon systems are comprised of parts that are subject to random failures. When a part fails, it must be replaced by an operable part that is provided by a supply network that supports the system. If replacement parts are not made available in a timely manner by the supply network, the weapon system becomes unavailable until replacements arrive.

Supply networks consist of many locations where spare parts are held, known as "echelons." Examples of echelons include depots, fulfillment centers, and customers. When many identical weapon systems operate in parallel and rely on a single supply network for replacement parts, decision makers must choose where and how to invest their resources into the purchase of spare parts in order to maximize the amount of time the systems are available for use. In particular, they must choose stock levels—the most efficient quantities of spare parts to stock at each echelon. Should decision makers choose stock levels that are too low, availability of the weapon systems suffers and unnecessary strain is placed on the supply network. Conversely, should decision makers choose stock levels that are too high, funds are wasted or inefficiently used (i.e., tied to parts in stock when they could be used for other purposes). As a result, decision makers face the problem of determining how to optimally choose stock levels. In particular, given a limited budget, a decision maker must determine how many spares of each part to buy, where in the supply network to stock these parts, and how much weapon system availability to expect.

To address these questions, this thesis formulates a mathematical model that determines the optimal way to invest limited budgets into the procurement of spare parts to maximize the expected availability of a weapon system. Specifically, we use stochastic optimization to maximize the expected number of time periods a weapon system is

available for use. The key decisions in our model determine the optimal stock levels of spare parts at each echelon in the system's supply network.

Our approach differs from other research in that most supply network optimization studies ignore interactions among part failures, repair, transit times, and costs. Consequently, decisions made to improve one component of the supply network may adversely impact other components. Without taking a comprehensive view of the system and making decisions accordingly, one cannot hope to achieve full efficiency.

The key contribution of our research is the integrated modeling of uncertainty, time, and monetary constraints on a multi-echelon supply network. We incorporate uncertainty into our modeling by randomly generating scenarios of time-based part failures for a particular weapon system. Then, we use those scenarios as input parameters for our stock level calculations while enforcing non-anticipativity. We also model the evolution of our inventory levels over time and incur a penalty in our objective function if the supply network does not provide replacement parts in a timely manner. Finally, we provide decision makers with an estimate of the weapon system availability they can expect given an optimal investment in spare parts for the system. To our knowledge, no other research incorporates these three aspects into a single mathematical model. This level of detail comes at the expense of lengthy solve times. Even for very small supply networks with few parts, the optimization solver requires a substantial amount of time to find an optimal solution. We view our research as the first step towards high-fidelity, multi-echelon modeling; perhaps future research can build upon our efforts and decrease the computational time required to solve our models.

2. Weapon System

Figure 1 illustrates a notional weapon system. The system is comprised of n different types of parts p_1, p_2,…, p_n. These parts operate in series, that is, the system is only available for use during a particular time period if all parts are working during that period. Each part has a known probability of failing after operating for a given number of periods. We assume that failures for each part are independent events (i.e., failure of one part does not impact the probability of failure of other parts).

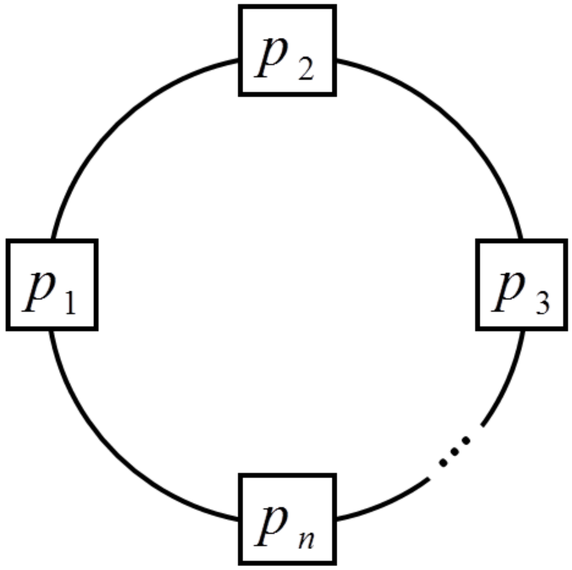


Figure 1. Notional weapon system.

a. Weapon System Availability

A weapon system is only valuable when it is available for operational use. This occurs when the quantity of working parts installed and/or spares on hand is at least the quantity required to operate the system. Accordingly, our research maximizes the availability of a weapon system over a defined time horizon by finding optimal stock levels of each part for each echelon in a system's supply network. We define *availability* of a weapon system as the fraction of time that system is available for use and calculate it as follows:

$$availability = \frac{t_{available}}{t_{available} + t_{not\ available}} \tag{1}$$

where $t_{available}$ and $t_{not\ available}$ are the number of periods the system is and is not available for use, respectively. By construction, the *availability* metric always produces a value between 0 and 1, providing us with an unambiguous measure with which to compare the effectiveness of different solutions.

b. *System of Systems Concept*

We further expand the scope of our *availability* metric to embody the "system of systems" concept, illustrated in Figure 2. We reason that many independent and identical weapon systems operate in parallel in one large, aggregate system. Accordingly, instead of measuring the *availability* of each individual system, our model makes stock level decisions that maximize *availability* of the aggregate system. This is defined as the average *availability* of the individual weapon systems that comprise the aggregate system. The system of systems concept applies to our problem in that each individual weapon system requires replacement parts from a single supply network that stocks spare parts for all weapon systems. When two weapon systems require the same part at the same time and the supply network only has one spare on hand, our model chooses to issue the spare to the entity that results in the largest average *availability*. If we considered each weapon system in isolation, we would neglect these types of interactions and improperly estimate the aggregate system's *availability*.

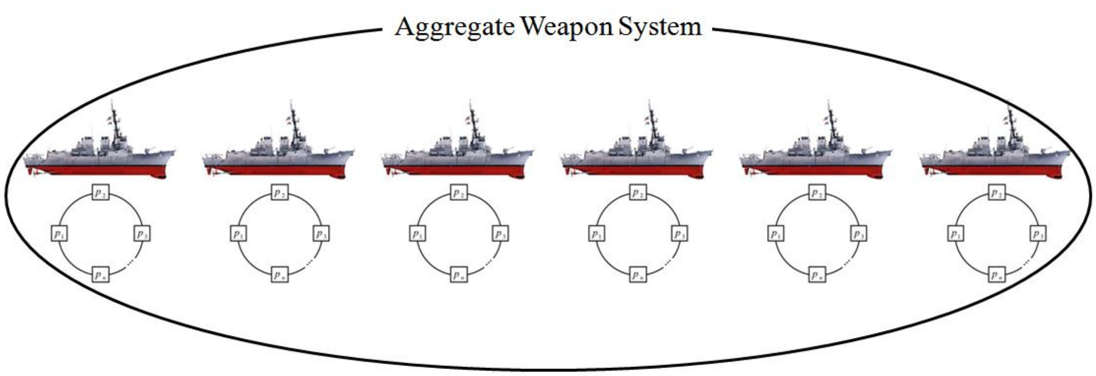


Figure 2. System of systems concept.

c. *Mapping Investments in Spare Parts to Availability*

A key benefit of our model is that it informs decision makers of how much weapon system *availability* they can expect, given monetary investments in spare parts for that system. Figure 3 illustrates a notional investment versus *availability* curve. The horizontal axis corresponds to our total monetary investments in spare parts for a particular weapon system and the vertical axis corresponds to the estimated fraction of

time a weapon system is available for use. Each point on the curve is a notional output of our model corresponding to the maximum average *availability* value that can be achieved by making optimal stock level decisions given a particular level of investment. For example, should a decision maker choose to invest $20 into the purchase of spare parts for the modeled weapon system, the curve shows that the weapon system is expected to be available 60 percent of the time. Note that initial investments in inventory result in large *availability* marginal gains. However, as monetary investments in spare parts increase, these gains decrease as diminishing returns take effect.

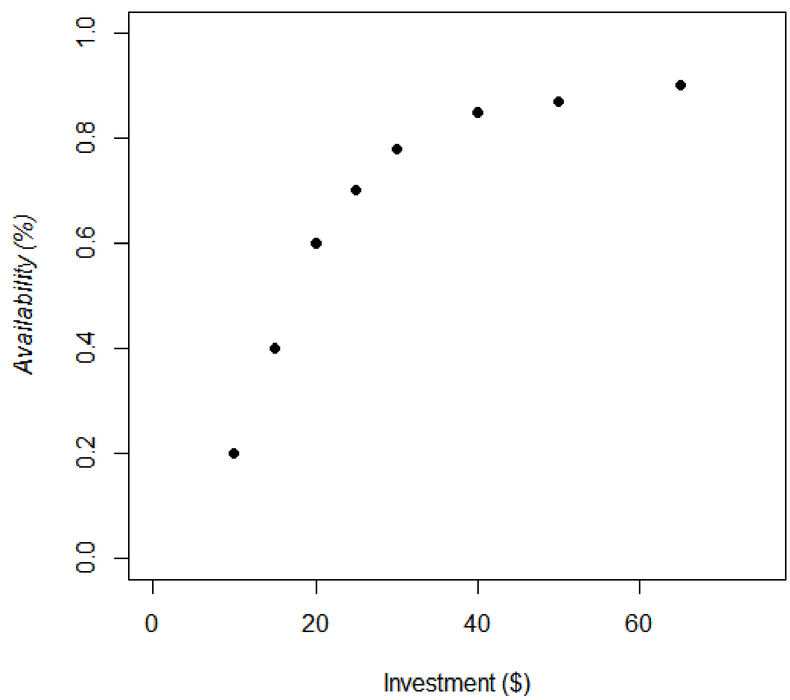


Figure 3. Notional curve of investment in spare parts versus *availability*.

3. Repairable Parts

In general, parts are categorized as either consumable or repairable. Typically, consumable parts are simple, low-cost items that are discarded after failure. Conversely, repairable parts are expensive, complex items that may be discarded after failure, but are

often repaired by a servicing echelon and placed back into circulation within a supply network. Entities that operate weapon systems comprised of repairable parts prefer purchasing repaired ones, because their price is often a fraction of what it would have cost to purchase the parts as new. A drawback to such a heavy reliance on repairable parts is that their availability within a supply network is often scarce due to long repair times. Since most weapon systems are comprised of repairable parts, our analysis focuses solely on these parts.

4. Multi-Echelon Supply Network

Figure 4 displays a simplified multi-echelon supply network for a repairable part p. Starting at the bottom of Figure 4, the first echelon in this network is referred to as the customer. The customer installs working parts to operate its weapon system and maintains a stock of ready-for-issue spares to replace installed parts that fail. Moving upstream from the customer, the second echelon is referred to as the fulfillment center. This echelon is located within the same geographic region as the customer, and it maintains a stock of spares that are used to supply replacement parts to the customer. Finally, moving upstream from the fulfillment center, the third echelon is referred to as the depot. The depot receives failed parts from the customer, repairs them, stores them, and issues them as replacement parts to supply the fulfillment center. All events take place within discrete time periods we denote as t.

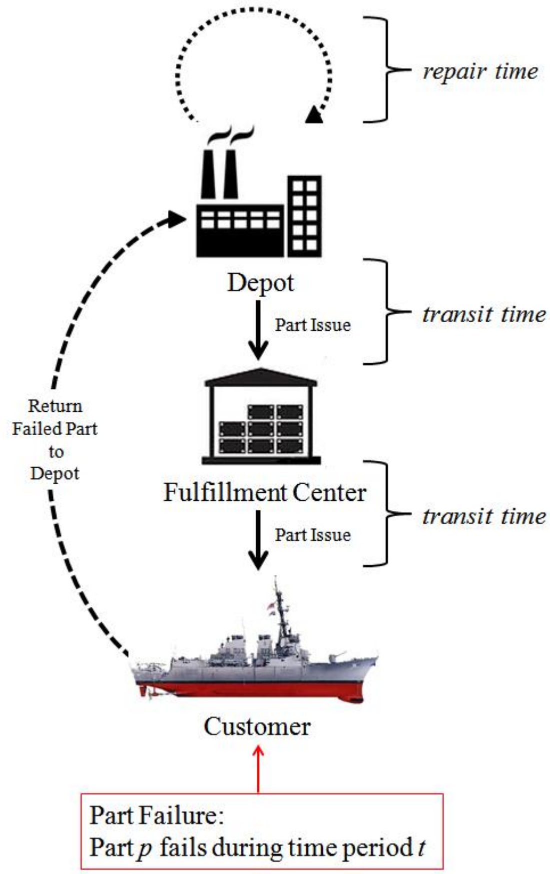


Figure 4. Basic multi-echelon supply network for repairable parts.

a. *Multi-Echelon Supply Network Events*

When a failure of part *p* occurs at customer echelon *e* during time period *t*, the following events take place:

- The customer checks to see if a spare *p* is on hand:
  - If so, the failed part is immediately replaced and the weapon system remains available for that time period. The customer's quantity of on-hand spares is decremented accordingly.
  - Otherwise, the weapon system becomes unavailable for the number of time periods that elapse until a replacement is received from the supporting fulfillment center.
  - In either case, the customer ships the failed part to the supporting depot for repair.
- If the customer requires a replacement part:

    - If the supporting fulfillment center has a spare on hand, it supplies the part to the customer and its quantity of on-hand spares is decremented accordingly. The time interval between the customer's part failure and the receipt of a replacement part from the supporting fulfillment center is denoted as transit_time.
    - Otherwise, the fulfillment center must await a replacement part from its supporting depot. The number of time periods it takes to transport a replacement part from the depot to the fulfillment center is also denoted as *transit_time*.
- When the depot receives a failed part from the customer, it repairs the part and makes it available for issue to any fulfillment center that requires the part. The number of time periods required to repair the part is denoted as *repair_time*.

b. Enhanced Multi-Echelon Supply Network

Adding to Figure 4, we enhance the scope of our multi-echelon supply network to take the shape of Figure 5. In this network, multiple depots supply replacement parts to multiple fulfillment centers, which, in turn, supply replacement parts to many customers. We assume that each depot specializes in repairing particular parts. Consequently, not all depots can repair all parts. Fulfillment centers, on the other hand, can receive, store, and issue all parts. Lateral supply only exists between fulfillment centers, and each customer is assumed to be supported by only one (pre-specified) fulfillment center.

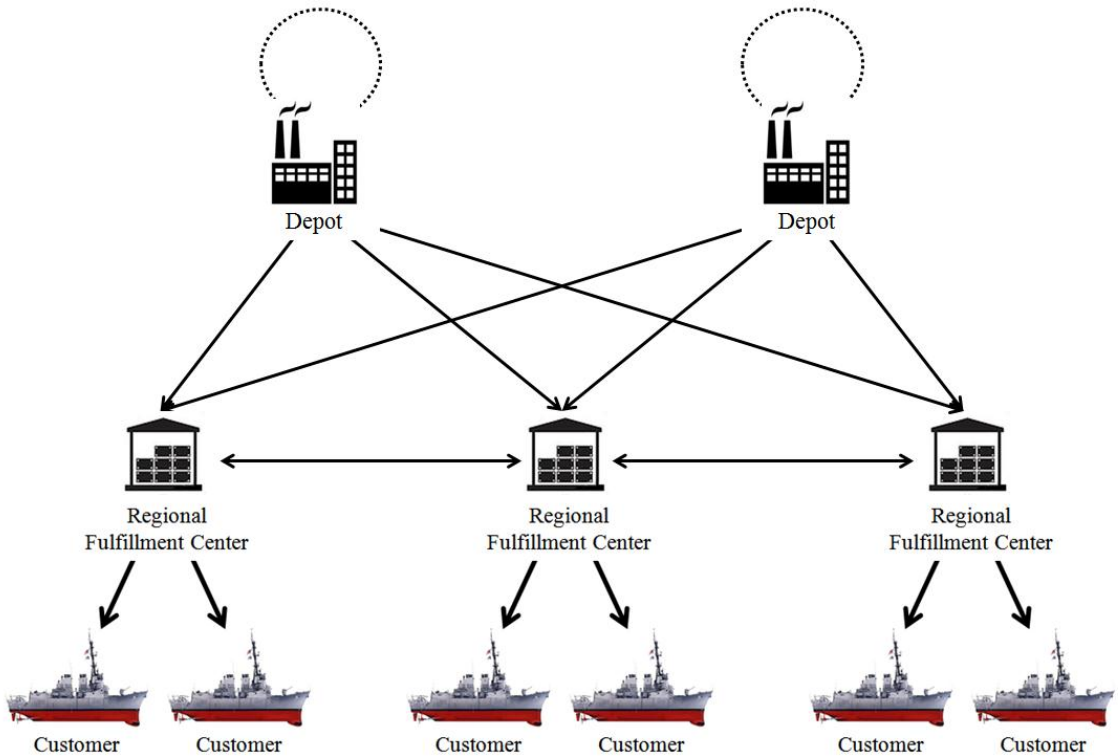


Figure 5. Enhanced multi-echelon supply network for repairable parts.

5. Budget

If our budget were unconstrained and we had unlimited storage space, we would choose to purchase infinite quantities of spare parts to store at each customer echelon. This would ensure that aggregate weapon system *availability* would always be 100 percent. Unfortunately, this is unrealistic. Therefore, we must constrain our model by restricting our monetary investments in spare parts (We ignore echelon capacity, which can easily be incorporated, if needed). Specifically, we include two constraints in our model, reflecting limits on our expenditures for the wholesale and retail echelons of the supply network. The wholesale budget represents the maximum amount of funding available to procure the initial quantity of spare parts to be stocked at depots and fulfillment center echelons. The retail budget represents the maximum amount of funding available to procure the initial quantity of spare parts to be stocked at all customer

echelons. We choose to model wholesale and retail budget constraints instead of a single, network-wide budget constraint so that our formulation aligns with real-world budgeting practices of weapon system support.

B. LITERATURE REVIEW

To establish a theoretical framework for our research, we build upon literature in the field of multi-echelon modeling. Work in this area dates back to the 1950s and has grown quite extensively in recent years due to advances in computing power. We categorize multi-echelon research into deterministic, stochastic, and simulation models, and borrow key ideas to guide our efforts.

1. Deterministic

We begin by examining research with deterministic parameters. Sherbrooke, often considered a pioneer in multi-echelon network modeling, formulates a nonlinear model that calculates stock levels by minimizing total backorders across all customers [1]. He uses a technique known as "marginal analysis" to arrive at optimal solutions. Muckstadt expands upon Sherbrooke's formulation and uses Lagrangian relaxation to minimize expected backorders subject to a constraint on inventory investment [2]. From both authors we gain an understanding of the link between weapon system *availability* and monetary investments in spare parts. Chandra develops a dynamic distribution model with warehouse and customer replenishment requirements [3]. His efforts focus on minimizing transportation, holding, delivery, and order costs in a network. From his research, we gain insight on how to formulate inventory balance constraints over a finite planning horizon and how to model the evolution of a supply network over time. Lee develops a continuous review, multi-echelon model for repairable items when emergency lateral transshipments are allowed between customer echelons [4]. His efforts expand on lateral shipment work previously done by Sherbrooke. Finally, Pirkul and Jayaraman use mixed-integer programming to minimize total transportation, distribution, and fixed costs for opening and operating plants and warehouses in a multi-commodity, tri-echelon network [5]. Customer locations and their demand for products are assumed to be known in advance.

2. Stochastic

To incorporate uncertainty into our modeling, we review research that incorporates stochastic processes. Schneider and Kirkpatrick introduce us to stochastic optimization and show us approaches for heuristic development [6]. Narrowing our focus to supply-specific research, we borrow ideas from Axaster on how to model lateral supply-determining when it is better for one or more supporting echelons to provide replacement parts to a supported echelon [7]. He develops decision rules for lateral supply in a single-echelon inventory system consisting of a number of parallel echelons with stochastic demand. His approach is to minimize expected costs. Graves determines inventory stock levels in a multi-echelon inventory system for repairable items [8]. His work presents an exact model for finding the steady-state distribution of net inventory levels for parts at each site as well as the distribution for the number of outstanding orders for a site at a particular time. From his work, we capture the events associated with modeling repairable parts.

Tsiakis et al. formulate a mixed-integer program that determines the number, location, and capacity of warehouses and distribution centers to be set up, the transportation links that need to be established in the network, and the flows and production rates of material, subject to uncertain demands [9]. They seek to minimize total annualized costs over the network. From their work, we gain an appreciation for approaches on how to handle uncertainty including the use of stochastically-generated scenarios. Moreover, we align our objective function with theirs, in that we want to find robust solutions that perform well over all scenarios of uncertain demand, not for a single presumed outcome. Caggiano et al. describe and validate a practical method for efficient computation of time-based service level requirements in a multi-item, multi-echelon service parts distribution system [10]. They use channel fill rates to determine where to position inventory in order to satisfy customer service agreements. Ganeshan formulates a multiple-retailer, single-warehouse, multiple-supplier model that encompasses both inventory and transportation factors [11]. Furthermore, his research finds near-optimal stocking policies for reorder points and order quantities under stochastic demands and lead-time constraints, subject to customer service constraints. From his work, we gather

ideas on the interactions between inventory and transportation decisions. Neale and Willems provide ideas on how to accommodate situations where the rate of demand changes over time [12]. Their formulation determines inventory locations and levels in a supply chain facing stochastic, non-stationary demand. Iida considers a periodic-review, dynamic, multi-echelon inventory problem, with non-stationary demands [13]. Moreover, his work expands our knowledge base by showing us the impacts of non-stationary demands over time. Simchi-Levi and Zhao provide methods for evaluating stochastic, multi-echelon inventory systems, specifically, the queuing inventory method, the lead-time-demand method, and the flow-unit method [14]. The queuing-inventory and lead-time-demand methods advise us of system performance measures at random points in time, whereas our approach more closely resembles the flow-unit method in that it models the impacts of decisions on time-based events. Ettl et al. formulate a constrained nonlinear optimization problem that minimizes the total average dollar value of inventory in a supply network, subject to meeting the service-level requirements of customers [15]. Finally, Acimovic and Graves provide us with ideas on how to enhance the functionality of our model for future research. Specifically, they formulate a model that minimizes average outbound shipping costs in a two-echelon network where business rules determine what fulfillment centers service which customers [16]. From their work, we gather ideas such as incorporating different service speeds to ship parts from one echelon to another, and the possible implementation of business rules to determine which echelon service customers require replacements.

3. Simulation

To round out our perspective on modeling approaches, we examine a study that uses simulation. Niranjan and Ciarallo formulate capacitated, three-echelon, and four-echelon systems with uncertainty in both demand and supply [17]. Their approach uses simulation-based optimization to determine optimal base stock levels for the components in the system, based on specified customer service levels. Their work gives us more insight on how to model the evolution of inventory levels over time.

Overall, our work differs from the cited literature in that our formulation comprehensively integrates uncertainty, time, and monetary constraints into a single, multi-echelon model, instead of looking at these aspects in isolation. Furthermore, we stray from the literature's heavy reliance on the use of expected values, in favor of modeling more detailed, high-fidelity parameters and variables that change over time.

C. PURPOSE

The purpose of our research is to develop an optimization model that provides guidance on the optimal investment in spare parts for a weapon system in order to maximize expected system *availability*. This includes modeling dynamic interactions of random part failures, time, and monetary constraints on the supply network. Even with very small networks, the number of stock level decisions that must be evaluated ventures into the tens of thousands; too many for a human to comprehend. Accordingly, we seek to leverage the objectivity and power of mathematical programming to help decision makers with this complex problem.

D. SCOPE, LIMITATIONS, AND ASSUMPTIONS

1. Scope

To scope our model formulation and subsequent analysis, we focus on three key areas. First, we only model repairable parts in a closed-supply network. This approach aligns with a conclusion drawn by Sherbrooke that repairable parts most directly affect a weapon system's *availability*, whereas, consumable parts do not [1]. Second, all of our echelons are in stationary locations. Third, we do not determine order sizes, reorder points, and safety stock levels with classic supply management formulas. Rather, we are concerned with calculating optimal stock levels that result in the best *availability* values over all randomly-generated part-failure scenarios.

2. Limitations

We limit the functionality of our model to only include the most essential features of a multi-echelon supply network. As formulated, the model may easily be modified to include many additional attributes such as more echelons and parts, implementing time-

varying transit times to ship parts from one echelon to another, and modifying the network so that different fulfillment centers support different customers according to business rules. Similarly, we have limited our experiments to only a few echelons and parts using notional data.

3. Assumptions

To shape the context of our problem and accommodate our limitations, the following assumptions apply to our supply network's policies and modeling logic.

a. Policies

- Our budget only pays for the initial number of parts stocked at each echelon. This excludes inventory management expenses such as ordering, holding, and transportation costs. Similarly, echelons incur no costs for the future receipt of replacement parts.
- All failed parts are repaired by depots; none are discarded.
- Depots immediately begin repair work on failed parts that they receive from customers.
- Repair and transit times are deterministic.
- Cannibalization of weapon systems is not allowed.
- Replacement parts for each echelon must be satisfied by parts circulating within the network. No external sources of supply exist.

b. Modeling Logic

- If a customer begins a time period with insufficient part(s) on hand, and receives the necessary quantity of replacements during the same period, the weapon system is considered available during that particular period.
- For the purposes of *availability*, part failures occur at the end of a time period.
- A shipment supplying a particular replacement part to a customer may be initiated during the same time period that the installed part fails.
- A maximum of one failure can occur per part, customer, and time period.

II. METHODOLOGY AND FORMULATION

A. METHODOLOGY

1. Map

Figure 6 portrays a high-level overview of our research methodology. For any weapon system we choose to analyze, we define the identified sets and parameters as inputs, and in return, our model delivers outputs that answer our research questions.

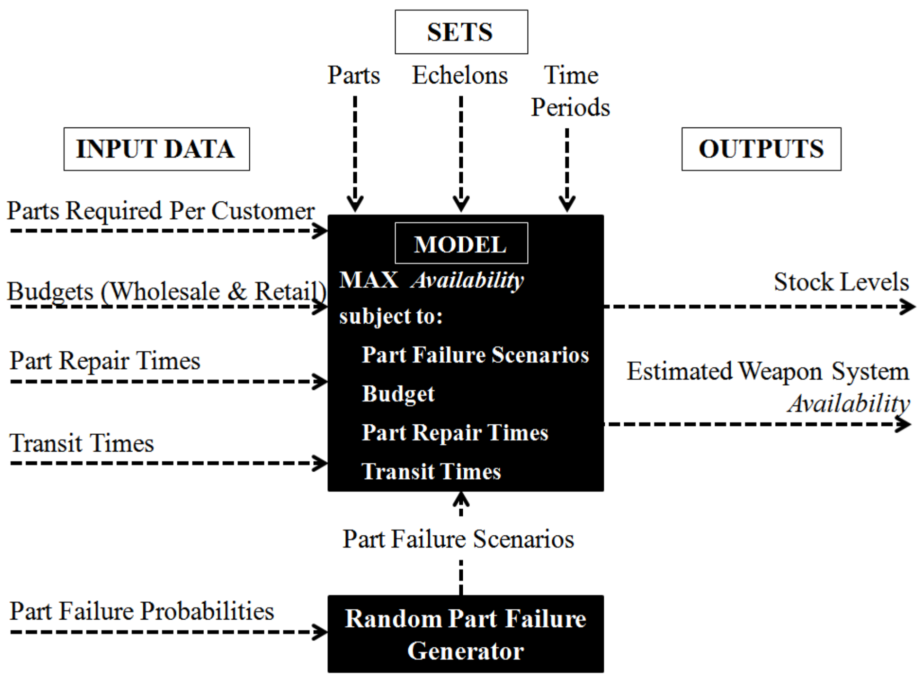


Figure 6. Methodology map.

2. Part Failures

Uncertainty within the supply network exists in many forms (e.g., part failure uncertainty, repair time uncertainty, and transportation time uncertainty). We focus our efforts on modeling part failure uncertainty, as part failures are considered the most uncontrollable and detrimental random events that occur within a supply network. When a part installed on a weapon system fails, the supply network must respond to this failure by making a replacement part available for issue as soon as possible, or else *availability*

of the weapon system suffers. We incorporate part failures into our formulation and represent the evolution of our network over time by modeling random inter-arrival times between part failures for each part installed at each customer. We do this by deriving part failure probability distributions from historical data and using a random part failure generator to create many randomly generated part-failure scenarios. Both are explained in the following section.

a. Part Failure Probability Distributions

Figure 7 displays example probability distributions of the operational time between failures for notional parts p_1 and p_2. The horizontal axes correspond to the number of time periods after replacement that each part has been operating and the vertical axes give the probability that each part fails after operating for that number of time periods. Both parts have a 100 percent probability of failing after operating for 35 time periods.

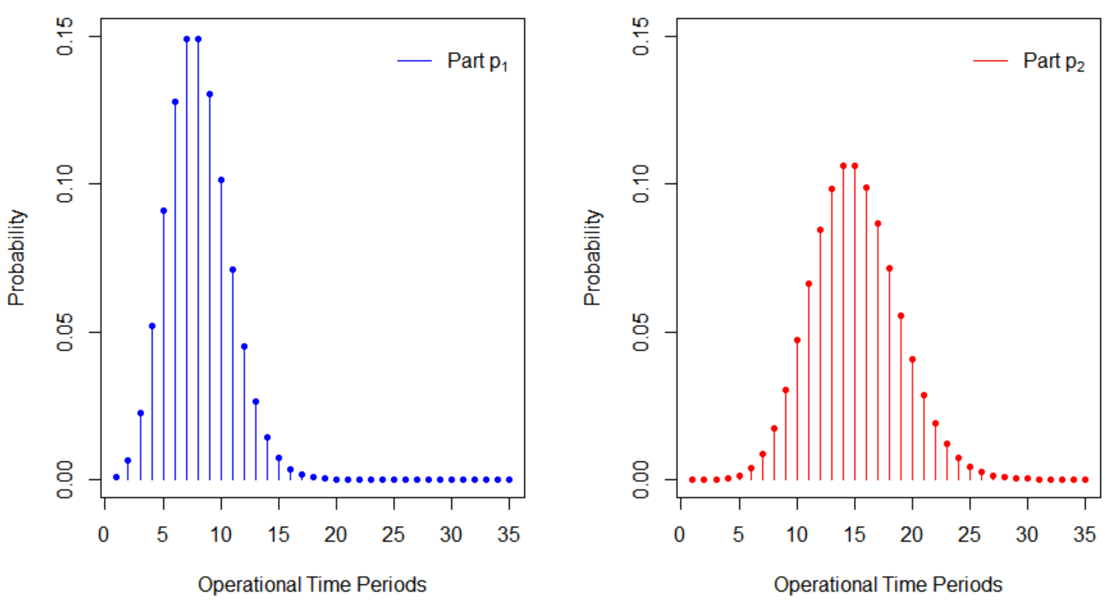


Figure 7. Probability distributions of the operational time between failures for notional parts p_1 (left) and p_2 (right).

b. Random Part Failure Generator

We have developed a part failure generator tool that accepts part-failure probability distributions as input and generates part-failure scenarios as output. In particular, the output of this tool is the parameter $fail\_time_{p,e,s,i}$, which describes the number of time periods during which echelon *e* has been operational when part *p* fails for the i^{th} time in scenario *s*. We use this tool to generate many scenarios of part failures, and in turn, use the scenarios as input parameters for our optimization model. Our model then employs robust optimization by calculating stock levels and expected *availability* over all input scenarios. This technique introduces variation into our metric calculations and ensures that we do not limit our results to deterministic solutions that are tied or subordinated to a specific scenario. The overall goal of our model is to choose stock levels for each echelon in such a way as to ensure that our supply chain performs well, on average, over all scenarios considered simultaneously.

3. Part Failures and *Availability*

Figure 8 illustrates the interactions between failures, transit times, and *availability*. Suppose we have a weapon system comprised of parts p_1 and p_2, and we seek to calculate the system's *availability* over the course of eight time periods. We assume that $transit\_time_{Fulfillment\ Center,Customer}$=2. For a particular scenario *s*, our random part failure generator calculates:

$$fail\_time_{p_1,Customer,s=1,i=1}=2$$

$$fail\_time_{p_2,Customer,s=1,i=1}=3$$

That is, part p_1 is scheduled to fail for the first time after the system has been operational for two time periods, and part p_2 is scheduled to fail for the first time after the system has been operational for three time periods. Accordingly, we commence operations and use the system during periods t_1 and t_2. At the end of period t_2, part p_1 fails. No replacements for part p_1 are on hand, so the system becomes unavailable for periods t_3 and t_4 while awaiting a replacement from the fulfillment center. At the

beginning of period t_5, a replacement for p_1 arrives and is immediately installed; thus, the system becomes available again during that period. At the end of period t_5, the system has been operational for three time periods, so part p_2 fails. Again, no replacements for part p_2 are on hand, so $transit\_time_{Fulfillment\ Center,Customer}$ forces the system to become unavailable for time periods t_6 and t_7. At the beginning of period t_8, a replacement for p_2 arrives and is immediately installed; thus, the system becomes available again during that period. Overall, the weapon system has been available for four time periods out of eight. Per Equation 1, the system's *availability* is 50 percent.

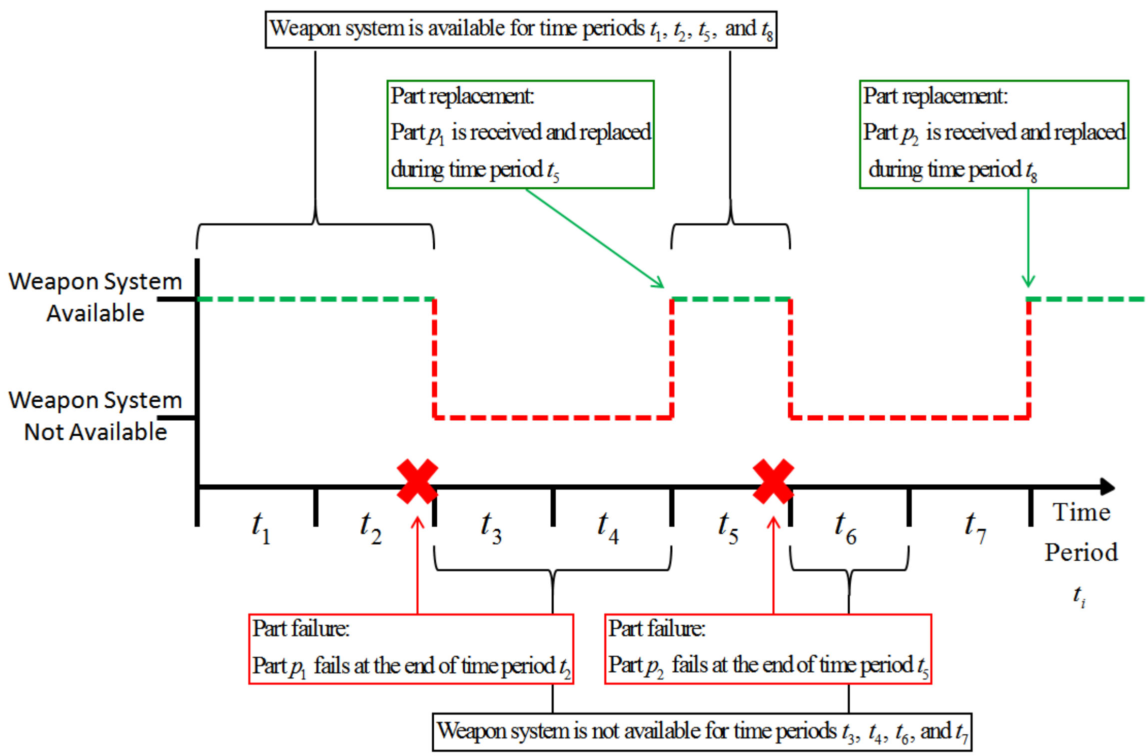


Figure 8. Illustration of weapon system *availability* calculation.

Note that of our part failure scenarios do not prescribe the actual time periods in which failures occur; rather, they indicate the number of *operational* time periods that elapse before a part fails. The operational status of the system depends, in turn, on our stock level decisions. For instance, if a spare of part p_1 had been present at time t_2, our

system would have remained operational during period t_3, and thus part p_2 would have failed at the end of period t_3, rather than at the end of period t_5. This subtlety introduces modeling challenges not present in most robust optimization problems, but it is necessary in order to accurately model the evolution of a weapon system.

B. FORMULATION

We now describe our mathematical model, which uses stochastic optimization to maximize aggregate weapon system *availability* by calculating optimal stock levels for spare parts in a multi-echelon supply network. It is important to note that our stock levels account for variation in the time of failures because they are chosen as the most efficient values considering our randomly-generated part failure scenarios. All other variables are specific to each scenario.

1. Sets and Indices

$p \in P$	part types
$e, e' \in E$	inventory echelons
$D \subseteq E$	depots
$F \subseteq E$	fulfillment centers
$C \subseteq E$	customers
$M \subseteq P \times E$	set of part-echelon pairs (p,e), where echelon e can process part p
$A \subseteq E \times E$	set of arcs (e,e') joining echelons e and e' in the shipping network
$t \in T$	time periods
$i \in I$	part failures
$s \in S$	randomly generated scenarios of part failures

2. Parameters [Units]

$repair\_time_{p,e}$	number of time periods required to repair part p at echelon e [time periods]
$part\_required_{p,e}$	minimum quantity of part p required to operate a weapon system at echelon e [parts]

$fail\_time_{p,e,s,i}$	operational time periods before part *p* fails at echelon *e* for the i^{th} time in scenario *s* [time periods]
$transit\_time_{e,e'}$	number of time periods required to ship a part from echelon *e* to echelon e' [time periods]
$cost_p$	cost of part *p* [$]
wholesale	budget to procure parts for depots and fulfillment centers [$]
retail	budget to procure parts for customers [$]
δ	relative weight on worst-case scenario of *availability* [unitless]

3. Decision Variables [Units]

$AVAILABLE_{e,t,s}$	binary; 1 if weapon system is available to operate at echelon *e* during time period *t* for scenario *s*, and 0 otherwise
$INV_{p,e,t,s}$	quantity of part *p* installed or on hand at echelon *e* during time period *t* for scenario *s* [parts]
$ENOUGH_{p,e,t,s}$	binary; 1 if $INV_{p,e,t,s}$ is greater than or equal to $part\_required_{p,e}$, and 0 otherwise
$FAIL_{p,e,t,i,s}$	binary; 1 if $fail\_time_{p,e,s,i}$ equals $\sum_{t' \le t} AVAILABLE_{e,t',s}$, and 0 otherwise
$SHIP_{p,e,e',t,s}$	quantity of part *p* shipped from echelon *e* to echelon e' during time period *t* for scenario *s* [parts]
$STOCK\_LEVEL_{p,e}$	maximum quantity of part *p* installed or on hand at echelon *e* at the beginning of the planning horizon [parts]
$WORST\_CASE$	worst-case system *availability* over all scenarios [fraction]

4. Equations

$$\underset{\substack{AVAILABLE,\, INV,\\ ENOUGH,\, FAIL, SHIP,\\ STOCK\_LEVEL,\, WORST\_CASE}}{\text{MAX}} \quad \frac{1-\delta}{|C|\cdot|T|\cdot|S|} \sum_{e \in C,t,s} AVAILABLE_{e,t,s} + \delta \cdot WORST\_CASE \qquad (2)$$

subject to:

$$WORST\_CASE \le \frac{1}{|C|\cdot|T|}\sum_{e\in C,t} AVAILABLE_{e,t,s} \qquad \forall s \tag{3}$$

$$INV_{p,e,t=1,s} = STOCK\_LEVEL_{p,e} \qquad \forall (p,e)\in M, s \tag{4}$$

$$INV_{p,e,t,s} = INV_{p,e,t-1,s} + \sum_{e'\in C, i | fail\_time_{p,e',s,i} \le t} FAIL_{p,e',t-repair\_time_{p,e}-1,i,s} - \sum_{\substack{e'|(e,e')\in A,\\ (p,e')\in M}} SHIP_{p,e,e',t,s} \tag{5}$$
$$\forall t\ge 2, (p,e)\in M, s \,|\, e\in D$$

$$INV_{p,e,t,s} = INV_{p,e,t-1,s} + \sum_{\substack{e'|(e',e)\in A,\\ (p,e')\in M}} SHIP_{p,e',e,t-transit\_time_{e',e},s} - \sum_{\substack{e'|(e,e')\in A,\\ (p,e')\in M}} SHIP_{p,e,e',t,s} \tag{6}$$
$$\forall t\ge 2, (p,e)\in M, s \,|\, e\in F$$

$$INV_{p,e,t,s} = INV_{p,e,t-1,s} + \sum_{\substack{e'|(e',e)\in A,\\ (p,e')\in M}} SHIP_{p,e',e,t-transit\_time_{e',e},s} - \sum_{i| fail\_time_{p,e,s,i} \le t-1} FAIL_{p,e,t-1,i,s} \tag{7}$$
$$\forall t\ge 2, (p,e)\in M, s \,|\, e\in C$$

$$\sum_{\substack{t'\le t, e'|(e',e)\in A,\\ (p,e')\in M}} SHIP_{p,e',e,t',s} \le \sum_{\substack{t'\le t,\\ i| fail\_time_{p,e,s,i} \le t'}} FAIL_{p,e,t',i,s} \tag{8}$$
$$\forall p, e\in C, t, s \,|\, (p,e)\in M, part\_required_{p,e} \ge 1$$

$$\sum_{\substack{t'\le t,\\ e'|(e',e)\in A,\\ (p,e')\in M}} SHIP_{p,e',e,t',s} \le \sum_{\substack{t'\le t,\\ e'\in C|(e,e')\in A,\\ i| fail\_time_{p,e',s,i} \le t'}} FAIL_{p,e',t',i,s} \tag{9}$$
$$\forall p, e\in F, t, s \,|\, (p,e)\in M$$

$$\sum_{\substack{t'\le t,\\ e'|(e,e')\in A,\\ (p,e')\in M}} SHIP_{p,e,e',t',s} \le \sum_{\substack{t'\le t, e'\in C,\\ i| fail\_time_{p,e',s,i} \le t'}} FAIL_{p,e',t',i,s} \tag{10}$$
$$\forall p, e\in D, t, s \,|\, (p,e)\in M$$

$$AVAILABLE_{e,t,s} \le ENOUGH_{p,e,t,s} \tag{11}$$
$$\forall p, e\in C, t, s \,|\, (p,e)\in M, part\_required_{p,e} \ge 1$$

$$AVAILABLE_{e,t,s} \ge \sum_{p|(p,e)\in M, part\_required_{p,e}\ge 1} ENOUGH_{p,e,t,s} - \left(\sum_{p|(p,e)\in M, part\_required_{p,e}\ge 1} 1\right) + 1 \tag{12}$$
$$\forall e\in C, t, s$$

$$ENOUGH_{p,e,t,s} \le \frac{INV_{p,e,t,s}}{part\_required_{p,e}} \qquad (13)$$
$$\forall p,e \in C, t, s \mid (p,e) \in M, part\_required_{p,e} \ge 1$$

$$ENOUGH_{p,e,t,s} \ge \frac{INV_{p,e,t,s} - part\_required_{p,e} + 0.99}{1 + part\_required_{p,e} + \left\lfloor \frac{retail}{cost_p} \right\rfloor} \qquad (14)$$
$$\forall p,e \in C, t, s \mid (p,e) \in M, part\_required_{p,e} \ge 1$$

$$FAIL_{p,e,t,i,s} \le 1 + \frac{\left(fail\_time_{p,e,s,i} - \sum_{t' \le t} AVAILABLE_{e,t',s} \right)}{t} \qquad (15)$$
$$\forall p,e \in C, t, i, s \mid (p,e) \in M, part\_required_{p,e} \ge 1, fail\_time_{p,e,s,i} \le t$$

$$FAIL_{p,e,t,i,s} \le \frac{\sum_{t' \le t} AVAILABLE_{e,t',s}}{fail\_time_{p,e,s,i}} \qquad (16)$$
$$\forall p,e \in C, t, i, s \mid (p,e) \in M, part\_required_{p,e} \ge 1, fail\_time_{p,e,s,i} \le t$$

$$\sum_{t \ge fail\_time_{p,e,s,i}} FAIL_{p,e,t,i,s} \le 1 \qquad \forall p,e,i,s \mid (p,e) \in M, part\_required_{p,e} \ge 1 \qquad (17)$$

$$\sum_{t' \mid fail\_time_{p,e,s,i} \le t' \le t} FAIL_{p,e,t',i,s} \ge \frac{1 - fail\_time_{p,e,s,i} + \sum_{t' \le t} AVAILABLE_{e,t',s}}{t - fail\_time_{p,e,s,i} + 1} \qquad (18)$$
$$\forall p,e,i,s \mid (p,e) \in M, part\_required_{p,e} \ge 1$$

$$INV_{p,e,t,s} \le STOCK\_LEVEL_{p,e} \qquad \forall p,e,t,s \mid (p,e) \in M, e \notin D \qquad (19)$$

$$STOCK\_LEVEL_{p,e} \ge part\_required_{p,e} \qquad (20)$$
$$\forall p,e \in C, s \mid (p,e) \in M, part\_required_{p,e} \ge 1$$

$$\sum_{(p,e) \in M \mid e \notin C} STOCK\_LEVEL_{p,e} cost_p \le wholesale \qquad (21)$$

$$\sum_{\substack{(p,e) \in M \mid e \in C, \\ part\_required_{p,e} \ge 1}} STOCK\_LEVEL_{p,e} cost_p \le retail + \sum_{\substack{(p,e) \in M \mid e \in C, \\ part\_required_{p,e} \ge 1}} part\_required_{p,e} cost_p \quad (22)$$

$$AVAILABLE_{e,t,s} \in \{0,1\} \qquad \forall e,t,s \tag{23}$$

$$ENOUGH_{p,e,t,s} \in \{0,1\} \qquad \forall p,e,t,s \tag{24}$$

$$INV_{p,e,t,s} \geq 0 \qquad \forall p,e,t,s \tag{25}$$

$$FAIL_{p,e,t,i,s} \in \{0,1\} \qquad \forall p,e,t,i,s \tag{26}$$

$$SHIP_{p,e,e',t,s} \geq 0 \qquad \forall p,e,t,s \tag{27}$$

$$STOCK\_LEVEL_{p,e} \geq 0 \qquad \forall p,e \tag{28}$$

$$WORST\_CASE \geq 0 \tag{29}$$

5. Objective Function

Our objective in Equation 2 is to maximize the weighted average of the *availability* of our weapon system over all part-failure scenarios and *availability* for the worst-case scenario. As calculated in Equation 3, use of $WORST\_CASE$ allows us to limit overestimation of *availability* over extreme values. In our computational examples, we typically set δ equal to 0.1.

6. Constraints

a. Inventory Levels

Equation 4 initializes inventory levels of each echelon in the first time period. In accordance with Figure 8, Equations 5 through 7 use $fail\_time_{p,e,s,i}$, $transit\_time_{e,e'}$, and $repair\_time_{p,e}$, to model the flow of parts through our supply network. Equation 5 calculates current inventory levels for depots as failed parts are received from customers, repaired, and issued to fulfillment centers as replacements. Similarly, Equation 6 calculates current inventory levels for fulfillment centers as parts are received from depots and subsequently issued as replacements to customers. Finally, Equation 7 calculates current inventory levels for customers as part failures occur and replacements are received from supporting fulfillment centers.

b. Shipping

Equations 8 through 10 prevent our model from anticipating and preemptively responding to part failures that occur in future time periods. This guarantees that prescient decisions do not falsely inflate our *availability* values based on knowledge we would not have in a real situation. Specifically, Equation 8 ensures that no customer receives more of part p than the total number of failures they have had for part p as of time t. Equation 9 specifies that no fulfillment center can receive more of part p than the sum of failures their customers have had for p as of time t. Finally, Equation 10 states that the total quantity of shipments of part p out of all depots cannot exceed the total number of failures of p over all customers as of time t.

c. Weapon System Availability

Equations 11 through 14 determine whether our weapon systems have sufficient quantities of each part on hand for each customer. Specifically, in Equation 13, if the current inventory of part p is less than the quantity of p required to operate the weapon system, the *ENOUGH* variable records this shortfall. Equations 11 and 12 ensure that the system becomes unavailable until enough parts are on hand.

d. Failure Events

Equations 15 through 18 signal the time periods when failures occur for each part in operation at each customer echelon. The complexity of these equations involves ensuring that each part fails in accordance with our randomly-generated part failure scenarios. We accomplish this with the parameter $fail\_time_{p,e,s,i}$ and the binary variables $AVAILABLE_{e,t,s}$ and $FAIL_{p,e,t,i,s}$. Equation 15 prevents part failures from occurring when the sum of operational time periods up to time t is strictly greater than $fail\_time_{p,e,s,i}$. Conversely, Equation 16 prevents part failures from occurring when the sum of operational time periods up to time t is strictly less than $fail\_time_{p,e,s,i}$. Equation 17 ensures that part p at echelon e cannot fail more than once for the i^{th} time over all time

periods in each scenario. Finally, Equation 18 forces $FAIL_{p,e,t,i,s}$ to signal a part failure when $fail\_time_{p,e,s,i}$ equals the sum of operational time periods up to time t.

e. Stock Levels

Equations 19 and 20 limit the values of our $STOCK\_LEVEL_{p,e}$ decision variable. Specifically, Equation 19 ensures that $INV_{p,e,t,s}$ never exceeds $STOCK\_LEVEL_{p,e}$, regardless of the scenario. Furthermore, $part\_required_{p,e}$ specifies the quantity of each part that must be on hand in order for a customer to operate its weapon system. Similar to Equation 19, we use Equation 20 to ensure that $STOCK\_LEVEL_{p,e}$ is always greater than or equal to this quantity.

f. Budget

Equations 21 and 22 set wholesale and retail budget limits.

THIS PAGE INTENTIONALLY LEFT BLANK

III. ANALYSIS

To conduct our analysis, we implement our formulation using the Generalized Algebraic Modeling System (GAMS) version 23.8.2 [18] and solve it as a mixed-integer linear program with the CPLEX 12.4.00 solver [19].

A. RESULTS

1. Investment versus *Availability*

To gain insights similar to those discussed in Section I.A.2.c, we temporarily simplify our formulation and consider a single budget constraint instead of separate wholesale and retail budgets. That is, we combine Equations 21 and 22 and limit the total cost to purchase spare parts for all echelons by a given budget value. Table 1 displays part attributes we use as model inputs for our experiments, and we model the flow of these parts throughout the supply network shown in Figure 10(d), over 15 time periods and 15 scenarios. Each customer requires one of each part to operate their weapon system, and our investment budget ranges from \$0 to \$100 in \$5 increments. Figure 9 illustrates the resulting investment versus *availability* curve.

Part Types	Average Failure Time Poisson Distribution (Time Periods)	$repair\_time_{p,Depot}$ (Time Periods)	$cost_p$ (\$)
p_1	1	1	6
p_2	2	2	1
p_3	3	3	3

Table 1. Repairable part attributes for experiments.

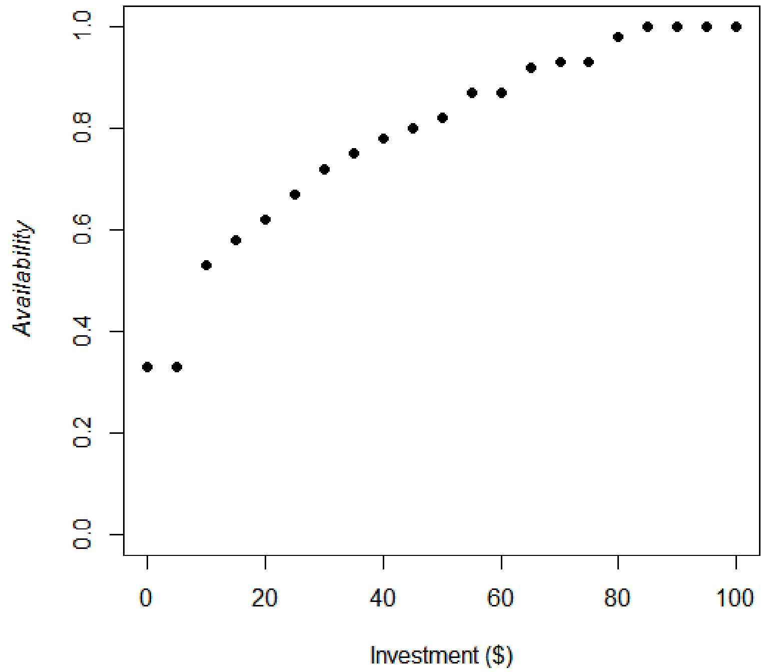


Figure 9. Investment versus *availability* curve.

Intuitively, our model produces results as expected. As our investment budget increases, our model chooses to purchase more spare parts to stock at echelons that result in the best *availability* values. Accordingly, we have more spares on hand to accommodate part failures, so *availability* ultimately improves. Clearly, our model exhibits the investment versus *availability* relationship discussed in Chapter I. Note that if we choose to invest $0 into purchasing spare parts, we estimate that our system will be available 33 percent of the time. Conversely, if we choose to invest $85 optimally, we estimate that our system will be available 100 percent of the time. Any investment beyond $85 does not increase *availability* for the scenarios considered.

Note that increasing the budget does not always result in *availability* gains. Recall that our formulation is based on the integrated modeling of uncertainty, time, and monetary constraints on a multi-echelon supply network's ability to support a weapon system. During any given time period, failures occur, parts are repaired, replacements are

in transit, and our model must choose the best stock levels that accommodate all of these events over all time periods and scenarios. Purchasing many parts of one type may not result in as great of an impact on *availability* as purchasing one of another type of part. For each \$5 increment in our investment budget, part types compete against each other. We purposely stagger our part costs from \$1 to \$6 to exaggerate this effect and ensure that our model cannot buy one more of each part with each budget increment. Thus, *availability* gains are not guaranteed each time the investment budget increases.

2. Multi-Echelon Supply Network Experiments

Next, we experiment with different sizes of multi-echelon supply networks to realize the impacts of topology and part types on *availability*. To do this, we reincorporate our wholesale and retail budgets and model part failure scenarios over each network shown in Figure 10. Table 2 itemizes initial conditions for our experiments. In order to reduce computational time, we increase the absolute tolerance (*availability* gap) for larger networks. Each customer is assumed to require one of each part to operate their weapon system.

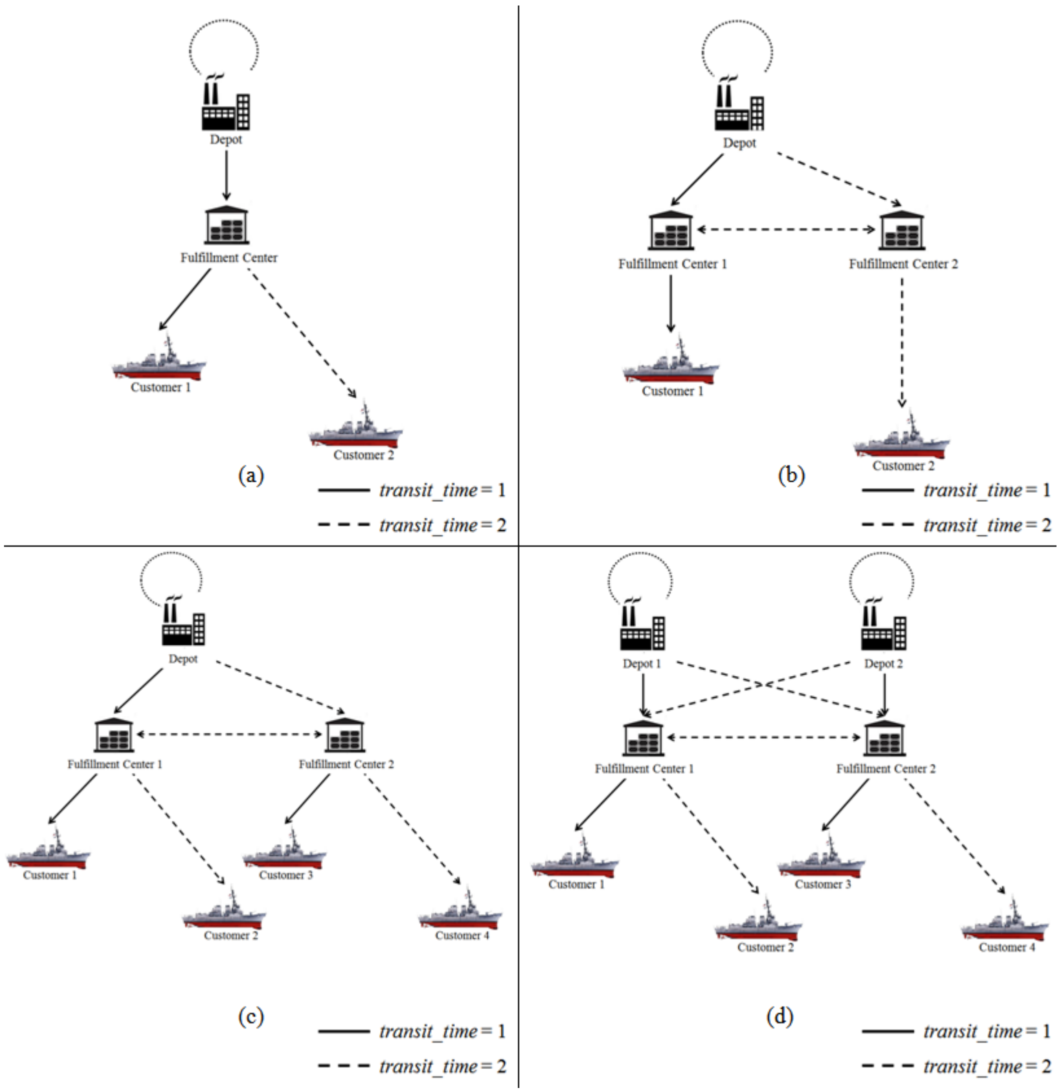


Figure 10. Multi-echelon supply networks for experiments.

Network	Topology	δ	Time Periods	Scenarios	Absolute Tolerance (%)
Four-Echelon	Figure 10(a)	0.1	15	15	5
Five-Echelon	Figure 10(b)	0.1	15	15	5
Seven-Echelon	Figure 10(c)	0.1	15	15	10
Eight-Echelon	Figure 10(d)	0.1	15	15	15

Table 2. Initial conditions for experiments.

Figure 11 displays our estimated *availability* values for each combination of wholesale and retail budgets for weapon systems comprised of parts p_1 and p_2. Likewise, Figure 12 displays *availability* for weapon systems comprised of parts p_1, p_2, and p_3. The horizontal axes indicate our retail budget and vertical axes indicate our wholesale budget. *Availability* values are shown at the intersection of each wholesale and retail budget combination.

Looking at Figure 11, we conclude that as supply network topology becomes more complex, more investments in spare parts are needed to ensure high rates of *availability*. Furthermore, we see that regardless of network size, if we choose to invest \$0 into purchasing spares for any echelon, our weapon system *availability* will never exceed 42 percent. The plots give us insight into the importance of having a sufficient retail budget. For example, if we choose to invest \$0 into the purchase of spares for all customers, no matter how much we invest into stocking spares at depots and fulfillment centers, we estimate *availability* will never exceed 79 percent. This is because weapon systems requiring replacement parts will always be unavailable for at least $transit\_time_{Fulfillment\ Center,Customer}$.

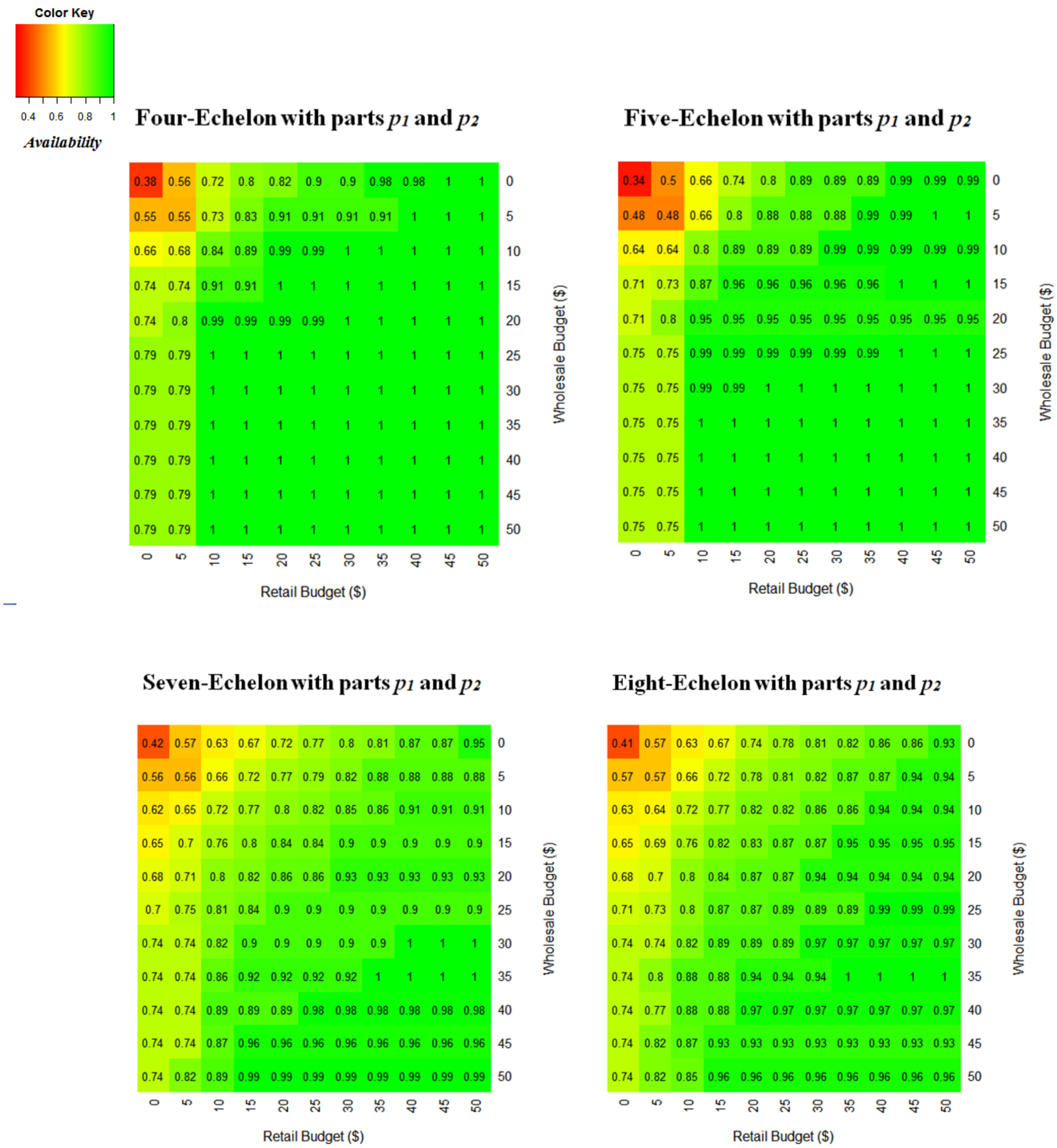


Figure 11. *Availability* estimates modeling parts p_1 and p_2.

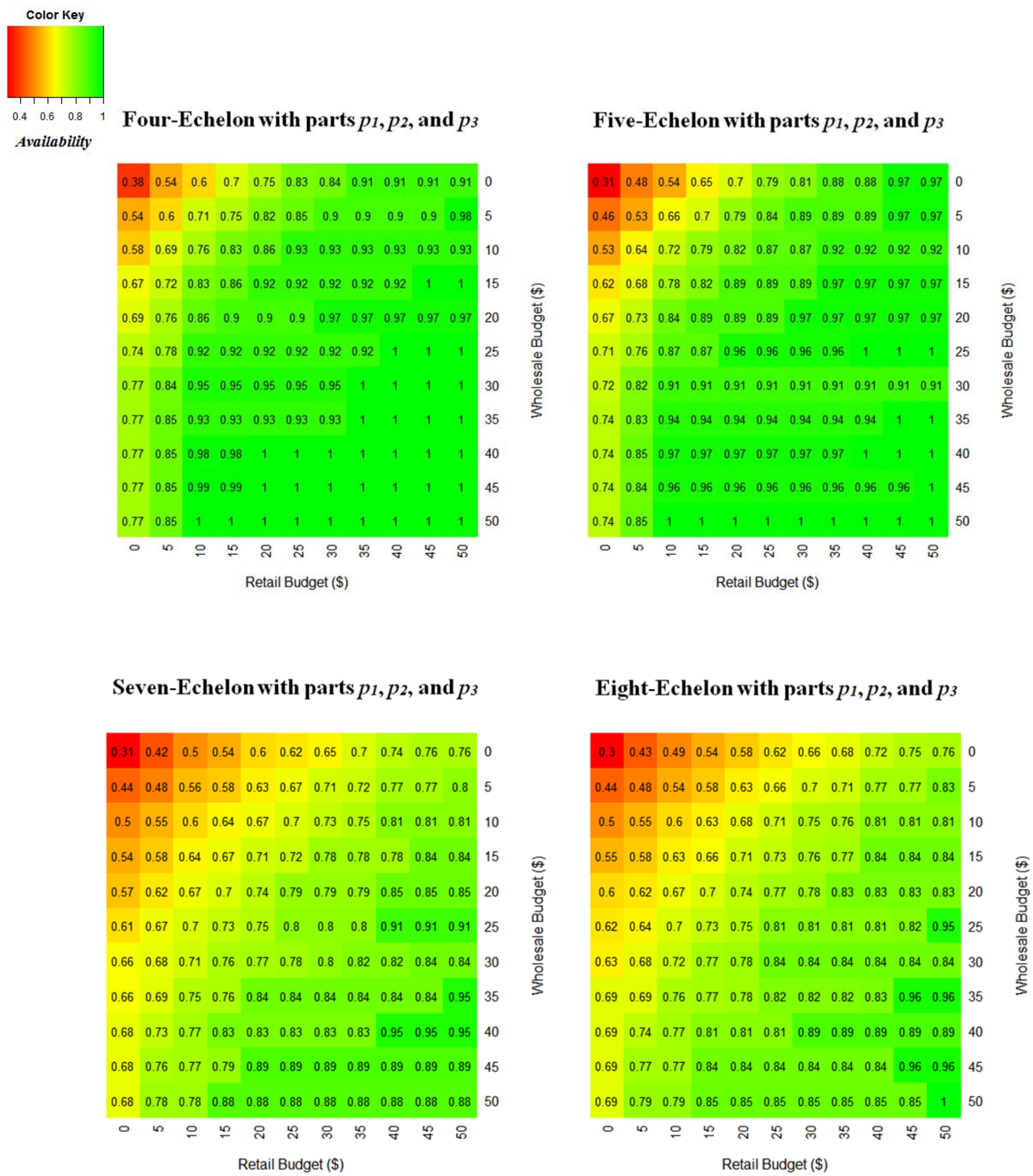


Figure 12. *Availability* estimates modeling parts p_1, p_2, and p_3.

Our experiments with three part types also show that, as the network becomes more complex, more investment is needed to achieve a certain *availability*. Moreover, regardless of network size, if we choose to invest \$0 into purchasing spares for any

echelon, at best, our weapon system *availability* will never exceed 38 percent. Comparing Figures 11 and 12, we see how modeling more part types results in visible trends of lower *availability*.

The general insight displayed by all of these networks is that for any fixed wholesale budget, *availability* increases at a faster rate when we increase our retail budget than when we increase our wholesale budget. We reason that this is most heavily influenced by $transit\_time_{Fulfillment\ Center, Customer}$. Recall that when customers have spare parts on hand, they can immediately replace failed parts and keep their weapon system operational, as opposed to waiting for replacement parts to arrive from a fulfillment center. Even though a customer's *availability* suffers while the part is in transit, a proper balance of spare part stock must be struck between depots and fulfillment centers because they have the ability to supply multiple customers.

To highlight efficiencies gained from multi-echelon modeling, we explore *availability* values from Figure 12(d), which corresponds to an eight-echelon network modeling the flow of parts p_1, p_2, and p_3. Suppose we wish to partition a given total budget optimally into wholesale and retail budgets. From Figure 12(d), we note that each partition may result in a different *availability* value. Figure 13 illustrates an investment versus *availability* curve where each data point represents the *availability* achieved for a particular partition of the budget shown on the horizontal axis.

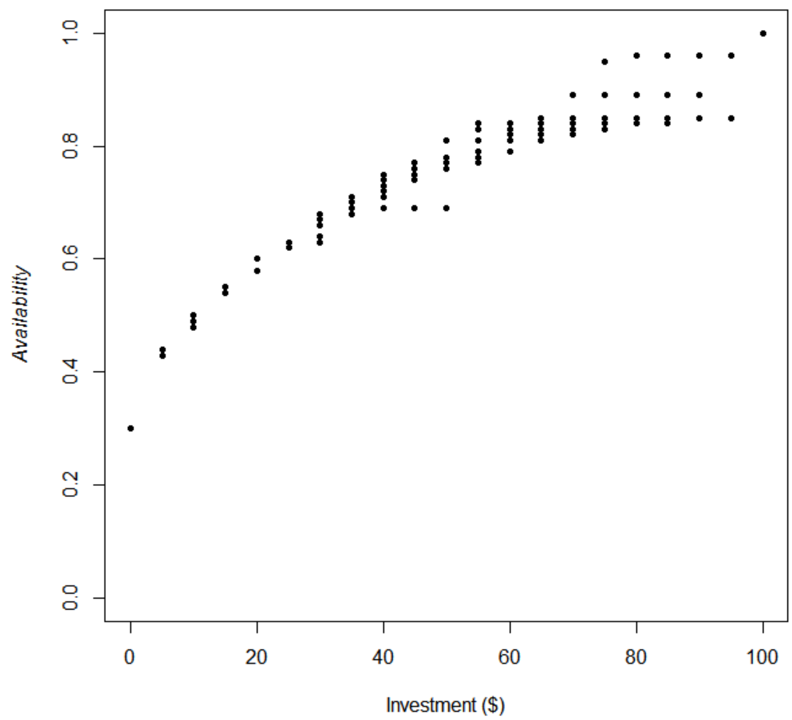


Figure 13. Investment versus *availability* curve for an eight-echelon network modeling parts p_1, p_2, and p_3. Each data point represents a particular division of the total budget indicated on the horizontal axis into wholesale and retail components.

Table 3 lists the particular *availability* values that we examine. From the data, we see that if a decision maker chooses to optimally invest any budget between $80 and $95 into purchasing spares, weapon system *availability* is expected to be 96 percent. Hence, a prudent choice is to only invest $80 and use the $15 savings elsewhere. Furthermore, if the decision maker chooses to accept a one percent reduction in *availability* from 96 percent to 95 percent, an additional savings of $5 can be realized. Investing $70 or less, results in a large drop in *availability*. We generalize that by taking advantage of efficiencies gained from multi-echelon modeling, for a smaller investment in spares, we may obtain the same or a comparable level of weapon system effectiveness.

Total Investment ($)	Wholesale Budget ($)	Retail Budget ($)	*Availability* (%)
95	45	50	96
90	45	45	96
85	35	50	96
80	35	45	96
75	25	50	95
70	40	30	89

$15 savings for same *Availability*

$20 savings if the decision maker accepts a 1% reduction in *Availability*

Table 3. *Availability* for an eight-echelon network modeling parts p_1, p_2, and p_3.

B. SOLVE TIME

Because we choose to model many of the detailed interactions that take place within a multi-echelon supply network, our formulation is challenged by lengthy solve time. To gain an appreciation for this, Table 4 displays the problem size of our experiments. Solve times for these instances are on the order of days. Clearly, even for our small experiments, finding an optimal solution is a substantial undertaking for the solver.

Network	Topology	δ	Part Types	Variables	Scenarios	Discrete Variables	Constraints
Four-Echelon	Figure 10(a)	0.1	p_1, p_2	7,399	15	5,328	15,167
		0.1	p_1, p_2, p_3	10,205	15	7,273	20,723
Five-Echelon	Figure 10(b)	0.1	p_1, p_2	9,141	15	6,678	18,317
		0.1	p_1, p_2, p_3	12,908	15	9,298	25,448
Seven-Echelon	Figure 10(c)	0.1	p_1, p_2	15,180	15	11,626	33,196
		0.1	p_1, p_2, p_3	20,990	15	14,394	45,207
Eight-Echelon	Figure 10(d)	0.1	p_1, p_2	15,180	15	10,992	33,196
		0.1	p_1, p_2, p_3	20,990	15	14,394	45,207

Table 4. Problem size of multi-echelon experiments.

To overcome this challenge, we examine our formulation for improvements that may reduce solve time. The user specifies most of the input data (e.g., part types, part failure probabilities, network topology, time periods, and budgets), so we do not consider modifying those items. However, an aspect that we can easily change is the number of

part-failure scenarios over which we optimize. Conceptually, the more scenarios we consider simultaneously, the more accurate our *availability* estimate will be; however, more scenarios result in longer solve times. Accordingly, we seek to determine an adequate number of scenarios that strikes a reasonable balance between *availability* precision and solve time.

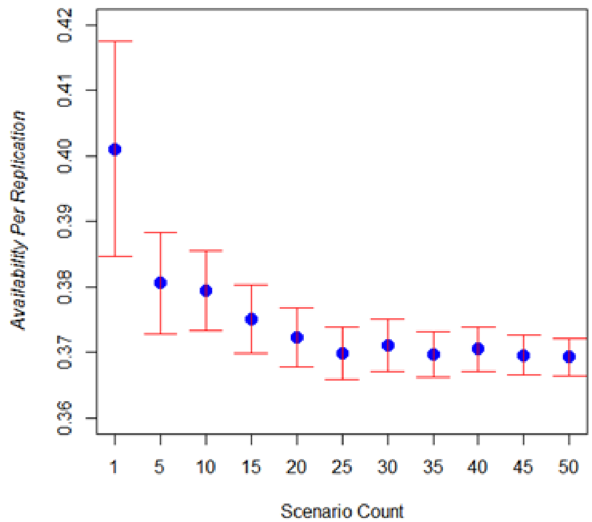
C. AVAILABILITY AND SOLVE TIME VARIATION

To gain an understanding of the variation in our output, we plot confidence intervals for *availability* and solve time with respect to the number of part failure scenarios over which we optimize. We accomplish this by conducting independent (i.e., different) replications for each experiment.

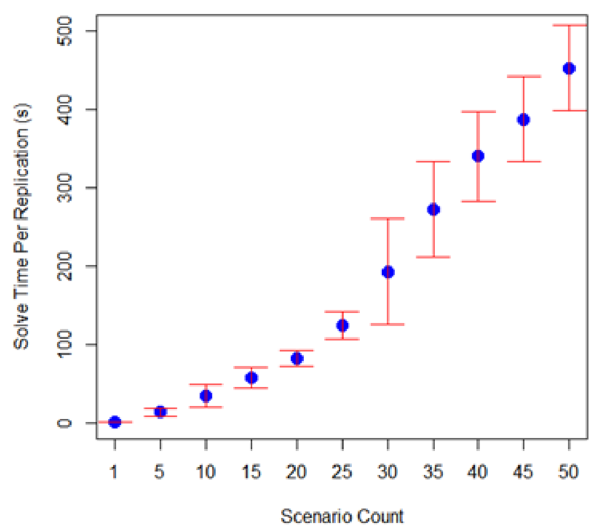
We perform this analysis with the four-echelon supply network shown in Figure 10(a) and model the flow of parts p_1 and p_2 over a horizon of 15 time periods. Table 5 displays part attributes. Wholesale and retail budgets are fixed at $5 each, and we solve to an absolute tolerance of five percent *availability*. Furthermore, we run 100 replications for each scenario, and our scenario counts range from 1 to 50 in increments of five. Figure 14 plots 95 percent confidence intervals for *availability* and solve time, given a particular number of scenarios for each replication. The horizontal axes indicate the number of part failure scenarios used within each replication and the vertical axes indicate *availability* and solve time, respectively.

Part Types	Average Failure Time Poisson Distribution (Time Periods)	$repair\_time_{p,Depot}$ (Time Periods)	$cost_p$ ($)
p_1	1	6	6
p_2	2	3	1

Table 5. Repairable part attributes for variation experiment.



(a)



(b)

Figure 14. 95 percent confidence intervals for *availability* and solve time.

With respect to *availability*, we note that as the number of part failure scenarios used increases from 1 to 50, our confidence intervals narrow and our *availability* estimates become more precise. Noticeably, if we choose to model only one scenario, we expect our results will have a large range of variation and our *availability* estimates will be overly optimistic (i.e., subordinating our stock level decisions to that solution will be highly suboptimal). The confidence interval for one scenario is over twice the width of

that of five scenarios, yet the computational time to solve is approximately the same. Thus, optimizing over multiple scenarios simultaneously is recommended. Our mean *availability* estimates and confidence interval widths reach a steady state when we employ 25 or more scenarios per replication. This informs us that modeling more than 25 scenarios would be an unnecessary use of computing resources.

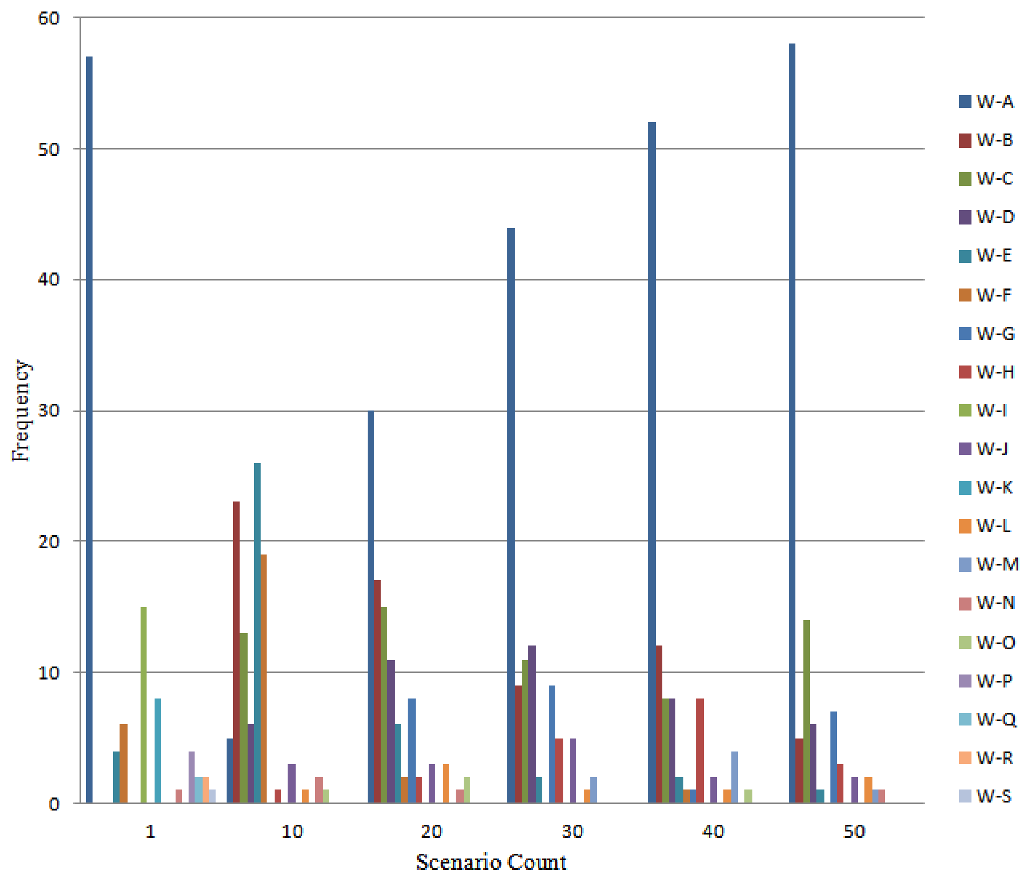
In stark contrast to our *availability* confidence interval plots, we see that the trajectory of the average solve time per replication resembles that of an exponential curve. Furthermore, as the scenario count per replication increases, so does the variation in the amount of time it takes to solve each replication. A drastic change in solve time variation occurs between 25 and 30 replications: the confidence interval for 30 scenarios per replication is nearly four times as wide as the interval for 25 scenarios per replication. Combining this observation with the fact that our *availability* confidence intervals reach a steady state at 25 scenarios per replication, we reason that modeling 25 scenarios per replication gives us outputs that reasonably balances *availability* precision and solve time for this particular experiment.

We generalize these two notions and conclude that for any particular supply network, part types, time horizon, and budget combination we choose to model, there exists a scenario count that strikes a reasonable empirical balance between *availability* precision and solve time, making the modeling of more scenarios unnecessary.

D. STOCK LEVEL CONVERGENCE

In addition to balancing *availability* precision and solve time, we predict that the number of part failure scenarios used also plays a role in our convergence to an optimal solution for $STOCK\_LEVEL_{p,e}$. Conceptually, we reason that as scenario count increases, our model should more frequently choose a specific stock level plan for each spare part at each echelon. To illustrate this concept, we conduct an experiment with the four-echelon supply network shown in Figure 10(a) and model the flow of parts p_1 and p_2 with wholesale and retail budgets fixed at \$5 each. We run 100 replications for each scenario where our scenario counts range from 1 to 50 in increments of five, and we solve to an absolute tolerance of one percent *availability*. Figures 15 and 16 display the

resulting wholesale and retail stock level histograms and corresponding plans. The horizontal axes indicate the number of part failure scenarios used within each replication and the vertical axes indicate the number of times a particular stock level plan is chosen as the optimal solution for each replication. Stock level plans accompany each histogram.

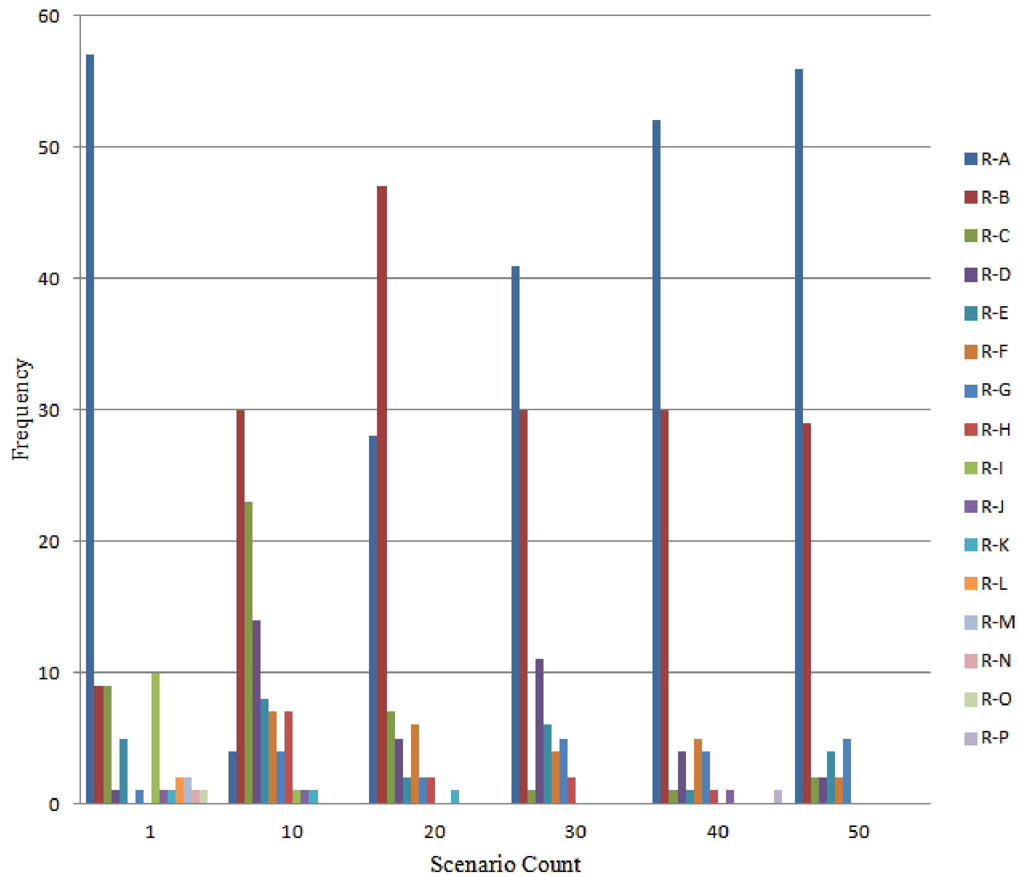


Stock Level Plan	W-A		W-B		W-C		W-D		W-E		W-F		W-G		W-H	
Type	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
Depot	0	3	0	0	0	1	0	0	0	1	0	0	0	1	0	0
Fulfillment Center	0	2	0	2	0	2	0	3	0	1	0	1	0	3	0	4

Stock Level Plan	W-I		W-J		W-K		W-L		W-M		W-N		W-O		W-P	
Type	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
Depot	0	0	0	0	0	1	0	2	0	1	0	2	0	2	0	5
Fulfillment Center	0	0	0	5	0	0	0	3	0	4	0	1	0	2	0	0

Stock Level Plan	W-Q		W-R		W-S	
Type	p_1	p_2	p_1	p_2	p_1	p_2
Depot	0	2	0	4	0	4
Fulfillment Center	0	0	0	1	0	0

Figure 15. Wholesale stock level histogram and plans.



Stock Level Plan	R-A		R-B		R-C		R-D		R-E		R-F		R-G		R-H	
Type	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
Customer 1	1	4	1	1	1	2	1	1	1	1	1	2	1	3	1	3
Customer 2	1	3	1	2	1	2	1	3	1	1	1	3	1	3	1	2

Stock Level Plan	R-I		R-J		R-K		R-L		R-M		R-N		R-O		R-P	
Type	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2	p_1	p_2
Customer 1	1	2	1	4	1	1	1	2	1	3	1	2	1	1	1	3
Customer 2	1	1	1	2	1	4	1	3	1	1	1	5	1	6	1	4

Figure 16. Retail stock level histogram and plans.

Per the figures, we can see that as scenario count increases, our model more frequently chooses a particular stock level plan (i.e., converges towards an optimal solution). Plans W-A and R-A are most frequently chosen as the plans that result in the best *availability* values when we optimize over one part failure scenario. However, recall that we anticipate it may be highly suboptimal to optimize over one scenario only; so, we focus our attention on results where scenario count is greater than one. Note that as we

increase the scenario count from 10 to 50, plans W-A and R-A steadily emerge as the most frequently chosen solutions while other stock level plans subside. Thus, we conclude that for this particular experiment, our model converges on stock level plans W-A and R-A as the optimal solutions for $STOCK\_LEVEL_{p,e}$. Note that for the experimental setup we have considered multiple optimal solutions are possible.

E. VALUE OF STOCHASTIC MODELING

To demonstrate the value of stochastic modeling, we now consider a simple system in which only 10 part failure scenarios are possible. In this simplified setting, we compare *availability* values resulting from optimization over a single scenario vice optimization over all 10 scenarios. We model the flow of parts p_1, p_2, and p_3 throughout the supply network shown in Figure 10(b) over 15 time periods and solve to an absolute tolerance of five percent *availability*. Table 6 displays part attributes for our comparisons, and wholesale and retail budgets are fixed at \$10 and \$15, respectively.

Part Types	Average Failure Time Poisson Distribution (Time Periods)	$repair\_time_{p,Depot}$ (Time Periods)	$cost_p$ (\$)
p_1	1	6	6
p_2	2	3	1
p_3	3	1	3

Table 6. Repairable part attributes for stochastic modeling experiment.

Let s and t denote part failure scenarios from the set $S = \{s_1, s_2, ..., s_{10}\}$. Furthermore, let $STOCK\_LEVEL_{p,e}^{s}$ denote the optimal stock level plan resulting from optimizing only over scenario *s*, and let $STOCK\_LEVEL_{p,e}^{ALL}$ denote the optimal stock level plan resulting from optimizing over all 10 scenarios in *S*. If we optimize over only a single scenario *s* and implement the resulting stock level plan, we cannot be certain that *s* is sufficiently representative of all possible scenarios. Should a different scenario occur, we may experience a substantially different *availability* value than anticipated, and one that is much lower than could have been achieved with perfect knowledge. To quantify

this difference, let $availability_{s,t}$ denote the *availability* value that results when we implement stock level plan $STOCK\_LEVEL^{t}_{p,e}$, but scenario *s* actually occurs. Likewise, let $availability_{s,ALL}$ denote the *availability* achieved when we implement $STOCK\_LEVEL^{ALL}_{p,e}$ and scenario *s* actually occurs.

Figure 17 displays the results of our ten-scenario experiment. The horizontal axis indicates the scenario *s* that actually occurred, while the data points indicate $availability_{s,t}$ for all $t \in S$, as well as $availability_{s,ALL}$. As the figure indicates, $availability_{s,ALL}$ is relatively high for all scenarios; in fact, for many scenarios, $availability_{s,ALL} = availability_{s,s}$. Furthermore, the risk of achieving a highly suboptimal *availability* value is much lower when using $STOCK\_LEVEL^{ALL}_{p,e}$. As the figure indicates, for a single outcome *s*, we see an absolute difference in *availability* of up to 30 percent depending on which scenario *t* was used in the optimization; in contrast, we have at most a 14 percent absolute difference between $availability_{s,ALL}$ and $availability_{s,s}$ for all scenarios. Table 7 summarizes the results of our experiment.

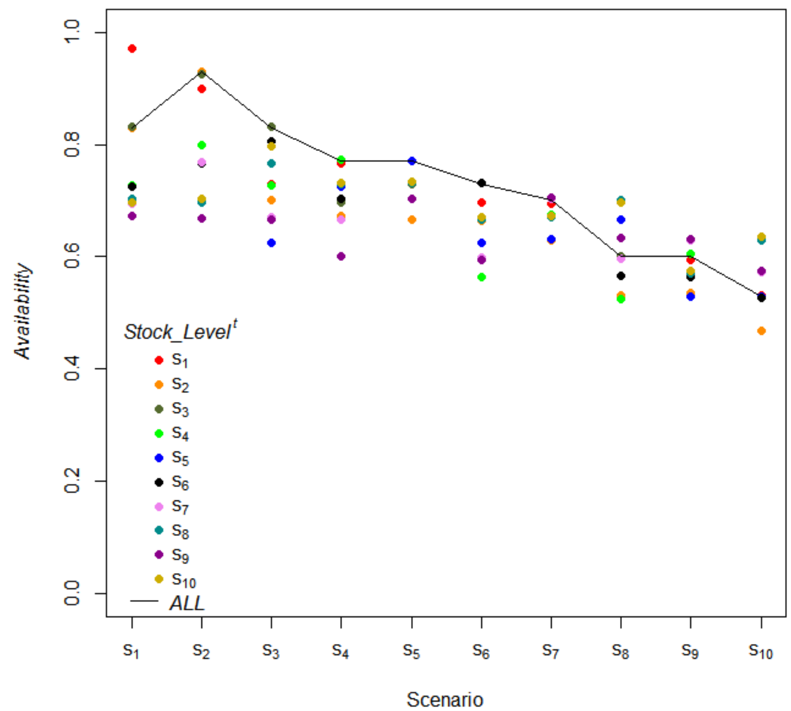


Figure 17. *Availability* plot for stochastic modeling experiment.

Scenario s	$availability_{s,s}$	Worst Case $availability_{s,t}$	Average $availability_{s,t}$	$availability_{s,ALL}$
s_1	0.97	0.67	0.75	0.83
s_2	0.93	0.67	0.79	0.93
s_3	0.83	0.63	0.73	0.83
s_4	0.77	0.60	0.71	0.77
s_5	0.77	0.67	0.72	0.77
s_6	0.73	0.57	0.66	0.73
s_7	0.70	0.63	0.67	0.70
s_8	0.70	0.53	0.61	0.60
s_9	0.63	0.53	0.58	0.60
s_{10}	0.63	0.47	0.55	0.53

Table 7. *Availability* results for stochastic modeling experiment.

F. TECHNIQUES TO REDUCE SOLVE TIME

Because our formulation models detailed interactions within a multi-echelon supply network, we face challenges in terms of solve time. We implement restriction and relaxation techniques to help our solver find an optimal solution faster.

For a restriction, we temporarily fix our $STOCK\_LEVEL_{p,e}$ decision variables to obtain an initial solution, and then use the corresponding variable bounds as an input, to obtain a full solution for our mixed-integer program. We accomplish this by fixing $STOCK\_LEVEL_{p,e}$ to equal the maximum quantity of spares we could purchase for each part at each echelon, given our wholesale and retail budgets. By solving this restriction, we obtain an initial integer solution for our problem, which occasionally helps the solver's optimization process.

In terms of implementing a relaxation, we suspend our shipping constraints to establish an upper bound on our optimal objective value. Then, we use this upper bound to incrementally refine bounds on our decision variables so that the solver more quickly converges on an optimal solution. We accomplish this by conducting a preliminary solve that excludes Equations 8 through 10. We reason that by suspending the shipping constraints, we create an environment where we foresee exactly when and where part failures will occur. In response, our model takes preemptive action and ships replacement parts to customers where failures are about to occur. Conceptually, our *availability* estimate should never be greater than in situations where we can see the future and take action accordingly. The discussed approaches result in modest reductions to solve time.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. SUMMARY AND FUTURE RESEARCH

A. SUMMARY

When many identical weapon systems operate in parallel and rely on a multi-echelon supply network for replacement parts, decision makers face the problem of determining how to optimally choose stock levels, given a limited budget.

Our research confronts this problem by formulating a mathematical model that determines the optimal investment of limited budgets into the procurement of spare parts so that weapon system *availability* is maximized. Specifically, we use stochastic optimization to integrate the modeling of uncertainty, time, and monetary constraints onto a multi-echelon supply network. First, we incorporate uncertainty into our model by randomly generating scenarios of time-based part failures (for a particular weapon system), which become input parameters for our stock level calculations. Second, we model the evolution of our inventory levels over time and incur a penalty in our objective function if the system is unavailable (i.e., if the supply network does not provide replacement parts in a timely manner). Accordingly, we provide decision makers with an estimate of expected *availability* given an optimal investment in spare parts for the system.

Rather than looking at supply network features such as part failures, repair, transit times, and costs in isolation, our formulation integrates their interactions in an optimization model that seeks maximizing overall expected *availability*. Our computational experiments show that, as supply network topology becomes more complex, more investments in spare parts are needed to ensure high rates of weapon system *availability*. We show the tradeoff relationship between solution accuracy and the time it takes to find an optimal solution as the number of scenarios considered increases.

B. FUTURE RESEARCH

We view our research as the first step towards detailed multi-echelon modeling that is not otherwise examined in the literature. Future research may build upon our work in terms of algorithmic refinements, technical enhancements, and further analysis.

1. Algorithmic Refinements

We propose that future work build upon our efforts to decrease solve time. Although we have performed a preliminary investigation of techniques to improve solve time, we believe that more can be done to improve the model's efficiency. Possible refinements may include reformulation of some constraints, especially those relating failures and *availability*, further experimentation with heuristic and exact solution methods, as well as decomposition approaches.

2. Technical Enhancements

Given that we seek to accurately estimate *availability*, we envision the addition of time-varying transit times and mobile customers as future technical enhancements to our model. As it stands, our model only accounts for part failure uncertainty and neglects any randomness associated with the number of time periods it takes to ship a replacement part from one echelon to another. In order to better estimate weapon system *availability*, we must account for this uncertainty. We foresee that the scope of our part failure scenarios should expand to include randomly generated transit times.

Additionally, in our current model, each customer is supported by the same fulfillment center over all time periods. This assumption oversimplifies reality because customers and their weapon systems are not stationary: customers may spend a certain time in close proximity to one fulfillment center and then move closer to another fulfillment center. Future work may model mobile customers by partitioning the network into regions and implementing a region index that specifies where particular events occur. Building on this, we recommend that future work formulates business rules that specify which fulfillment centers provide replacement parts to which customer for each time period.

3. Further Analysis

In terms of further analysis, we foresee great utility in applying our formulation to larger multi-echelon supply networks. As it stands, our analysis only considers the notional experiments discussed in Chapter III. Real-world weapon systems and supply

networks may encompass hundreds of echelons, thousands of repairable parts, and millions of investment dollars, far beyond what our research examines. Future work may model spare-part management of an actual weapon system from which additional operational insight can be gained.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] C. C. Sherbrooke, “Introduction,” in *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques*, F. S. Hillier, Ed., 2nd ed. Boston, MA: Kluwer Academic Publishers, 2004, pp. 3–10.

[2] J. A. Muckstadt, “A continuous time, multi-echelon, multi-item system with time-based service level constraints,” in *Analysis and Algorithms for Service Parts Supply Chains*, T. V. Mikosch, S. M. Robinson, and S. I. Resnick, Eds. New York: Springer, 2005, pp. 109–138.

[3] P. Chandra, “A dynamic distribution model with warehouse and customer replenishment requirements,” *J. Oper. Res. Soc.*, vol. 44, no. 7, pp. 681–692, Jul. 1993.

[4] H. L. Lee, “A Multi-echelon inventory model for repairable items with emergency lateral transshipments,” *Management Science*, vol. 33, no. 10, pp. 1302–1316, Oct. 1987.

[5] H. Pirkul and V. Jayaraman, “Production, transportation, and distribution planning in a multi-commodity tri-echelon system,” *Transportation Science*, vol. 30, no. 4, pp. 291–302, Nov. 1996.

[6] J. Schneider and S. Kirkpatrick, “Overview of optimization heuristics,” in *Stochastic Optimization.* Leipzig, Germany: Springer Berlin Heidelberg, 2006, pp. 43–47.

[7] S. Axsater, “A new decision rule for lateral transshipments in inventory systems,” *Management Science*, vol. 49, no. 9, pp. 1168–1179, Sep. 2003.

[8] S. C. Graves, “A multi-echelon inventory model for a repairable item with one-for-one replenishment,” *Management Science*, vol. 31, no. 10, pp. 1247–1256, Oct., 1985.

[9] P. Tsiakis, N. Shah, and C. C. Pantelides, “Design of multi-echelon supply chain networks under demand uncertainty,” *Ind. Eng. Chem. Res*., vol. 40, pp. 3585–3604, Jul. 2001.

[10] K. E. Caggiano, P. L. Jackson, J. A. Muckstadt, and J. A. Rappold, “Efficient computation of time-based customer service levels in a multi-item, multi-echelon supply chain: A practical approach for inventory optimization,” *Eur. J. Oper. Res*., vol. 199, pp. 744–749, Dec. 2009.

[11] R. Ganeshan, “Managing supply chain inventories: A multiple retailer, one warehouse, multiple supplier model,” *Int. J. Prod. Econ.*, vol. 59, pp. 341–354, 1999.

[12] J. J. Neale and S. P. Willems, “Managing inventory in supply chains with nonstationary demand,” *Interfaces*, vol. 39, no. 5, pp. 388–399, Sep.–Oct. 2009.

[13] T. Iida, “The infinite horizon non-stationary stochastic multi-echelon inventory problem and near-myopic policies,” *Eur. J. Oper. Res*., vol. 134, pp. 525–539, Nov., 2001.

[14] D. Simchi-Levi and Y. Zhao, “Three generic methods for evaluating stochastic multi-echelon inventory systems,” unpublished.

[15] M. Ettl, G. E. Feigin, G. Y. Lin, and D. D. Yao, “A supply network model with base-stock control and service requirements,” *Oper. Res*., vol. 48, no. 2, pp. 216–232, Mar.–Apr., 2000.

[16] J. Acimovic and S. Graves, “Making better fulfillment decisions on the fly in an online retail environment,” unpublished.

[17] S. Niranjan and F. W. Ciarallo, “Base-stock levels in multi-echelon inventory systems with multiple intermediary product demands,” in *Industrial Engineering Research Conf.*, Nashville, TN, 2007, pp. 298–303.

[18] GAMS. 2012. General Algebraic Modeling System. [Online]. Available: http://www.gams.com/download

[19] GAMS. 2012. CPLEX. [Online]. Available: http://www.gams.com/dd/docs/solvers/cplex.pdf

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California